

# Studying chemical reactions in biological systems with MBN Explorer: implementation of molecular mechanics with dynamical topology

Gennady B. Sushko<sup>1,2</sup>, Ilia A. Solov'yov<sup>3,4</sup>, Alexey V. Verkhovtsev<sup>1,2</sup>, Sergey N. Volkov<sup>5</sup>, and Andrey V. Solov'yov<sup>2,4</sup>

<sup>1</sup> Goethe-Universität Frankfurt am Main, Max-von-Laue-Str. 1, 60438 Frankfurt am Main, Germany

<sup>2</sup> MBN Research Center, Altenhöferallee 3, 60438 Frankfurt am Main, Germany

<sup>3</sup> University of Southern Denmark (SDU), Campusvej 55, 5230, Odense M, Denmark

<sup>4</sup> On leave from A.F. Ioffe Physical-Technical Institute, Polytekhnicheskaya ul. 26, 194021 St. Petersburg, Russia

<sup>5</sup> Bogolyubov Institute for Theoretical Physics, Metrolohichna str. 14-b, Kiev, 03680, Ukraine

**Abstract.** The concept of molecular mechanics force field has been widely accepted nowadays for studying various processes in biomolecular systems. In this paper, we suggest a modification for the standard CHARMM force field that permits simulations of systems with dynamically changing molecular topologies. The implementation of the modified force field was carried out in the popular program MBN EXPLORER, and, to support the development, we provide several illustrative case studies where dynamical topology is necessary. In particular, it is shown that the modified molecular mechanics force field can be applied for studying processes where rupture of chemical bonds plays an essential role, e.g., in irradiation- or collision-induced damage, and also in transformation and fragmentation processes involving biomolecular systems.

## 1 Introduction

Nowadays, it has become feasible to study structure and dynamics of molecular systems that constitute of millions of atoms [1, 2] and evolve on time scales up to hundreds of nanoseconds [3] by employing the classical molecular mechanics (MM) approach. In this approach, a molecular system is treated classically, so that constituent atoms interact with each other through a parametric phenomenological potential that is governed by the type of individual atoms and by the network of chemical bonds between them. This network defines a so-called molecular topology, that is a set of rules that impose constraints in the system and permit maintaining its natural shape, mechanical, and thermodynamical properties. The MM method has been widely used throughout the last decades [2, 4–6] and implemented in the well-established computational packages, such as CHARMM [7], AMBER [8], GROMACS [9], and NAMD [10].

MBN EXPLORER ([www.mbnexplorer.com](http://www.mbnexplorer.com)) [11] is an alternative, emerging software for the simulation of complex biomolecular, nano- and mesoscopic systems. It is suitable for classical molecular dynamics (MD), Monte Carlo [12–16] and relativistic dynamics simulations [17–20] of a large range

of molecular systems, such as nano- [21, 22] and biological systems, nanostructured materials [23, 24], composite/hybrid materials [25–28], gases, liquids, solids and various interfaces [29, 30], with the sizes ranging from atomic to mesoscopic. Among other applications, MBN EXPLORER can be used to simulate thermo-mechanical damage of a biological medium, e.g. a DNA nucleosome, which is caused by the propagation of a shock wave initiated by irradiation with fast ions [31]. The results of such simulations are used then to evaluate the efficiency of radiation with different projectiles [32] within the framework of the multiscale approach to the physics of radiation damage [33] and can be applied in the field of ion-beam cancer therapy [32–35].

Despite numerous successes, the conventional MM method is primarily capable of studying processes where chemical reactions do not take place. This leads to significant limitations of the method and makes it practically unsuitable for studying highly non-equilibrium processes in biomolecular systems, e.g. thermo-mechanical biodamage. This particular example involves rupture and formation of covalent bonds that cannot be simulated by the conventional MM method due to a fixed topology of the system.

Simulation of the rupture and formation of covalent bonds can be performed by using Quantum Mechanical/Molecular Mechanical (QM/MM) methods or *ab initio* MD simulations [36–38]. Both methods are computationally rather demanding and, thus, the *ab initio* approach is used typically for studying fragmentation of small biomolecules, such as DNA nucleobases or nucleotides [39, 40]. The size of such systems is far from the typical sizes of systems of biological relevance, consisting of hundred thousands of atoms, and more. This problem is addressed to some extent in QM/MM methods where a core part of a large biomolecular system is described quantum mechanically while all the surroundings are described classically using, for example, the conventional MM method [41, 42]. Thus, the rupture or formation of covalent bonds can be simulated only in a small part of the system, which is treated quantum mechanically.

In this paper, we present an extended version of the conventional MM method, which has been recently implemented in MBN EXPLORER, and demonstrate that this extension describes correctly the dynamically changing molecular topology of a system within the classical MD framework. The presented modification takes into account additional parameters of the system, such as dissociation energy of bonds, bonds multiplicity and the valence of atoms. The functional form of the interatomic interactions is also adjusted to account for the finite dissociation energy of the chemical bonds. Finally, three examples of simulations with the extended MM method are presented to demonstrate a proof of principle for utilizing the force field with dynamic topology. The first two case studies illustrate the processes of rupture and formation of covalent bonds in a small biomolecule, namely an alanine dipeptide, which is one of the simplest building blocks of larger biomolecular systems like polypeptides or proteins. Having proven the force field to work on a simple dipeptide allows us to generalize the framework towards macromolecules. This will allow for studying the systems of biologically relevant sizes, on the time scales which are not accessible by means of *ab initio* methods. The last example illustrates the process of water splitting and the evolution of chemical equilibrium. It is demonstrated that the results of the simulation are in a reasonable quantitative agreement with those of the analytical calculations.

## 2 Theoretical approach

### 2.1 Molecular mechanics potential

The MM potential represents a phenomenological parametrization of the potential energy of a system and is widely used to describe structure and prop-

erties of macromolecular systems, such as polypeptides [43–47], proteins [47–51], DNA [32, 52–54], lipids [51, 55–57], and many others. All physically important interactions in a system, i.e. both covalent and non-bonded long-range interactions, are accounted for using a simple parametric form, so that the total energy of the system reads as:

$$U_{\text{tot}} = U_{\text{cov}} + U_{\text{vdW}} + U_{\text{Coul}}. \quad (1)$$

The terms on the right-hand side describe the covalent, van der Waals, and electrostatic Coulomb interactions, respectively. The van der Waals interaction between two neutral atoms or molecules is usually modeled by the Lennard-Jones (LJ) potential:

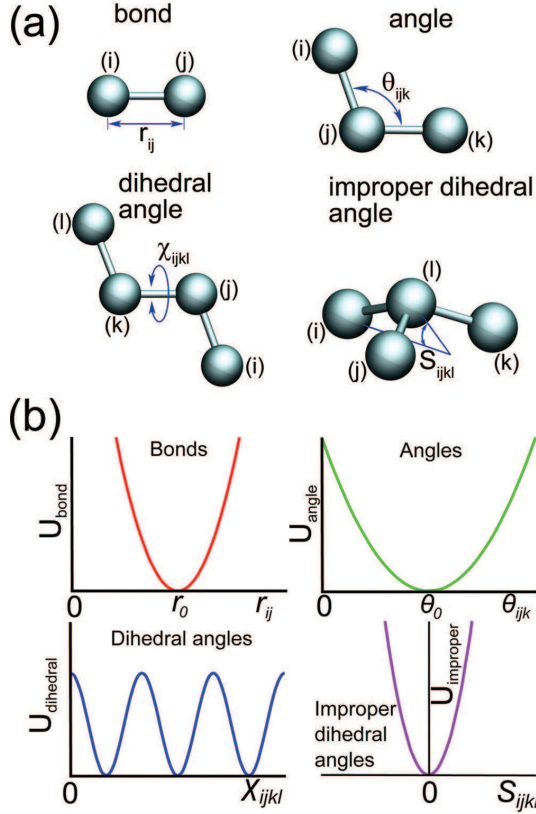
$$U_{\text{LJ}} = \epsilon \left[ \left( \frac{r_0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_0}{r_{ij}} \right)^6 \right], \quad (2)$$

where  $\epsilon$  is the depth of the potential energy well,  $r_0$  is the equilibrium distance, and  $r_{ij}$  is the distance between atoms. The function  $U_{\text{cov}}$  parameterizes the covalent interactions through a number of empirical parameters and fitting functions. MBN EXPLORER implies the following parametrization for the  $U_{\text{cov}}$  term:

$$U_{\text{cov}} = \sum_{\substack{\alpha=1 \\ i,j \in \alpha}}^{N_b} U_{ij}^{(\text{bond})} + \sum_{\substack{\alpha=1 \\ i,j,k \in \alpha}}^{N_a} U_{ijk}^{(\text{angle})} + \sum_{\substack{\alpha=1 \\ i,j,k,l \in \alpha}}^{N_d} U_{ijkl}^{(\text{dihedral})} + \sum_{\substack{\alpha=1 \\ i,j,k,l \in \alpha}}^{N_i} U_{ijkl}^{(\text{improper})}. \quad (3)$$

It is important to stress here that MM potential is only applicable to the systems with a predefined topology, i.e. a set of rules that define chemical bonds in the system. The energy of the covalent interactions is then modeled as a sum over all such interactions. The first and the second terms on the right-hand side of Eq. (3) describe the components of the potential energy of the system arising due to stretching of the bonds between two atoms and due to variation of angles between every topologically defined triplet of atoms, respectively (see Figure 1a). The third term is known as the torsion energy, which is characterized through a dihedral angle formed by every four atoms connected via covalent chemical bonds. The last term describes the so-called improper dihedral angles that are used in the molecular topology to maintain planarity.

A widely used MM potential, called CHARMM [58], represents a set of parameters for the simulation of structure and dynamics of bio-/macromolecular systems. This force field employs harmonic approximation for describing the interatomic interactions, thereby limiting its applicability to small deformations of the molecular system. The form of the potential energy functions, which are



**Fig. 1.** (a) Internal coordinates describing molecular mechanics interactions:  $r_{ij}$  governs bond stretching,  $\theta_{ijk}$  represents the angular term,  $\chi_{ijkl}$  gives the dihedral angle, and the small out-of-plane angle  $S_{ijkl}$  is governed by the so-called improper dihedral angle. (b) Dependencies of the potential energy on coordinates used in the molecular mechanics potential, Eq. (3), describing the bonded, angular, dihedral angular, and improper dihedral angular interactions.

used to describe the components of  $U_{\text{cov}}$  in the standard CHARMM force field, is demonstrated in Figure 1b. In the case of substantial deformations, the interaction forces should decrease to zero as the valence bonds rupture. The rupture of valence bonds should also cause the involved angular and dihedral interactions to vanish as well. In this work, we report on advances in the MBN EXPLORER development, that permit classical MD simulations of the rupture of covalent bonds by using a dissociative CHARMM potential. This approach goes beyond the harmonic approximation, thus describing the physics of molecular dissociation more accurately, and permits construction of dynamic molecular topology, which instructs MBN EXPLORER how the existing covalent bonds can break and new covalent bonds can be formed. These features make MBN EXPLORER rather unique, for example, for simulating irradiation- and collision-induced bio-damage by means of classical MD. To the best of our knowledge, presently there is no similar ap-

proach implemented in other software for classical MD simulations and, therefore, the implemented algorithms are seen as unique know-how of MBN EXPLORER. The adopted methodology is described in detail further in this section.

## 2.2 Rupture of covalent bonds

In order to model rupture of covalent bonds in the CHARMM force field, MBN EXPLORER uses modified interaction potentials which describe the interactions of atoms connected by chemical bonds. The standard CHARMM force field treats the covalent interactions within the harmonic approximation as

$$U_{ij}^{(\text{bond})} = k_{ij}^b (r_{ij} - r_0)^2. \quad (4)$$

Here  $k_{ij}^b$  is the force constant of the bond stretching,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , and the parameter  $r_0$  is the equilibrium covalent bond length. The above parametrization describes well the bond stretching regime in the case of small deviations from  $r_0$  but gives an erroneous result for larger distortions. For a satisfactory description of the covalent bond rupture it is reasonable to substitute the parabolic potential with the Morse potential. It requires one additional parameter, as compared to the harmonic approximation (4), and takes into account the energy of bond dissociation. Potential energy of the system of two atoms interacting via the Morse potential reads as:

$$U_M(r_{ij}) = D_{ij} \left[ e^{-2\beta_{ij}(r_{ij}-r_0)} - 2e^{-\beta_{ij}(r_{ij}-r_0)} \right], \quad (5)$$

where  $D_{ij}$  is the bond dissociation energy and the parameter  $\beta_{ij}$  determines steepness of the potential. It follows from Eq. (5) that  $U_M(r_0) = -D_{ij}$ .

Let us consider a small deformation of the covalent bond from its equilibrium distance,  $r_{ij} - r_0 \ll r_0$ . In this case, the energy of the bond can be approximated harmonically as:

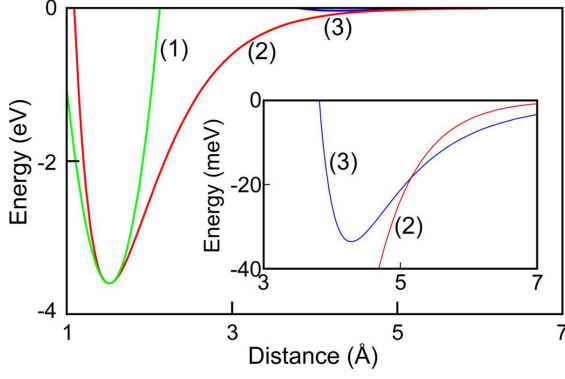
$$U_M(r_{ij}) \approx -D_{ij} + \beta_{ij}^2 D_{ij} (r_{ij} - r_0)^2 k, \quad (6)$$

so that

$$\beta_{ij} = \sqrt{k_{ij}^b / D_{ij}}. \quad (7)$$

This expression defines  $\beta_{ij}$  for a certain covalent bond and relates it to the value of  $k_{ij}^b$  used in the standard CHARMM force field.

Figure 2 illustrates the Morse potential which models the interaction of two carbon atoms, having the CN7 – CN8B type according to the CHARMM nomenclature. This covalent bond occurs, for example, between the C4'-C5' atoms of ribose. At small deviations from  $r_0$ , the Morse potential (red curve) and the harmonic approximation (green curve) are close to each other. With increasing interatomic



**Fig. 2.** The pairwise carbon-carbon (CHARMM type CN7 – CN8B) interaction potential in harmonic (1) and Morse approximations (2). The van der Waals interaction between the two atoms is illustrated with the blue line denoted as (3).

distance ( $r_{ij} \gg r_0$ ), the atoms start to interact through polarization forces modeled through the Lennard-Jones potential. The comparison between the Morse and the Lennard-Jones potentials at larger distances is shown in the inset. It follows that both potentials are close to each other at the distances of about the van der Waals contact distance for the non-bonded interactions, which is about 2 Å for the considered carbon atoms.

### 2.3 Rupture of valence angles

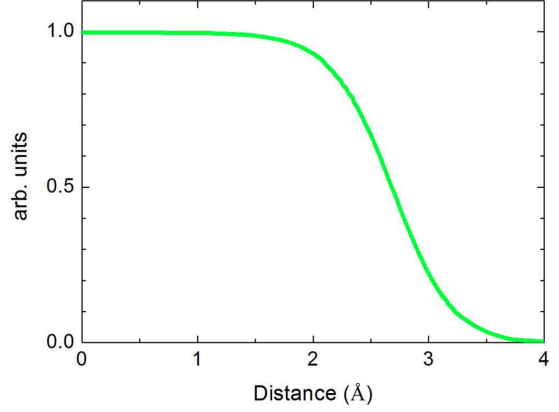
The rupture of chemical bonds in the course of simulation automatically employs an improved potential for the valence angles. In the CHARMM force field, the potential associated with the change of a valence angle between bonds with indices  $ij$  and  $jk$  reads as:

$$U_{ijk}^{(\text{angle})} = k_{ijk}^a (\theta_{ijk} - \theta_0)^2, \quad (8)$$

where  $k_{ijk}^a$  and  $\theta_0$  are parameters of the potential, and  $\theta_{ijk}$  is the actual value of the angle formed by the three atoms. This potential grows rapidly with increasing the angle, and it may lead to non-physical results when modeling the covalent bond rupture. In order to avoid such cases, in the modified force field the harmonic potential (8) is substituted with an alternative parametrization, which reads as:

$$U_{ijk}^{(\cos)} = 2k_{ijk}^a [1 - \cos(\theta_{ijk} - \theta_0)]. \quad (9)$$

At small variations of the valence angle, this parametrization is identical to the harmonic approximation (8) used in the standard CHARMM force field. For larger values of the angle, the new parametrization (9) defines an energy threshold which becomes important for an accurate modeling of bond breakage.



**Fig. 3.** The switching function  $\sigma(r_{ij})$  calculated for the carbon-carbon (CHARMM type CA –CA) interaction using Eq. (10). The function is used to rupture angular interactions in the modified CHARMM force field.

The rupture of a covalent bond is accompanied by a rupture of the angular interactions associated with this bond. The effect of bond breakage on the angular potential can be described through a special function  $\sigma(r_{ij})$ , defined as

$$\sigma(r_{ij}) = \frac{1}{2} \{1 - \tanh [\beta_{ij}(r_{ij} - r_{ij}^*)]\}, \quad (10)$$

with  $r_{ij}^* = (R_{ij}^{\text{vdW}} + r_0)/2$ . Figure 3 illustrates that  $\sigma(r_{ij})$  has a form of a smoothed step function. This function introduces a correction to the angular interaction potential, assuming that the distance between two atoms involved in an angular interaction increases from the equilibrium value,  $r_0$ , up to the van der Waals contact value,  $R_{ij}^{\text{vdW}}$ . Since an angular interaction depends on two bonds connecting the atoms with indices  $ij$  and  $ik$ , the potential energy, describing the valence angular interaction that is subject to rupture, is parameterized as

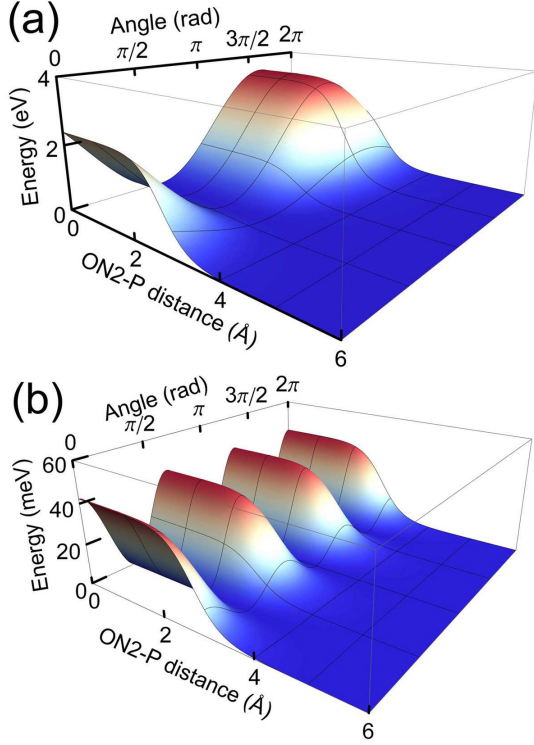
$$\tilde{U}_{ijk}^{(\text{angle})} = \sigma(r_{ij}) \sigma(r_{jk}) U_{ijk}^{(\cos)}. \quad (11)$$

As seen from this expression, the angular potential decreases with the increase of the bond length between any of the two pairs of atoms  $ij$  or  $jk$ . For exemplary purposes, in Figure 4a we show the CN8B – ON2 – P angular potential, which arises, for instance, when modeling DNA nucleotides. The presented angular potential was calculated using Eq. (11) assuming the breakage of the bond between the oxygen and the phosphorous atoms. For the sake of illustration, the CN8B – ON2 bond length in this case was taken equal to its equilibrium value  $r_0$ .

### 2.4 Rupture of dihedral interactions

“Dihedral” interactions arise in the conventional MM potential due to the change of the dihedral angles between every four topologically defined atoms.





**Fig. 4.** (a) The CN8B – ON2 – P angular potential calculated using Eq. (11) with account for the ON2 – P bond rupture. (b) The CN4 – P – ON2 – CN7 dihedral potential calculated using Eq. (13) with account for the ON2 – P bond rupture.

Let us consider a quadruple of atoms with indices  $i, j, k$  and  $l$  (see Figure 1), bound through an interaction which is governed by a change of the dihedral angle. In this case, the dihedral angle stands for the angle between the plane, formed by the atoms  $i, j$  and  $k$ , and the plane, formed by the atoms  $j, k$  and  $l$ . In the harmonic approximation, the dihedral-energy contribution reads as:

$$U_{ijkl}^{(\text{dihedral})} = k_{ijkl}^d [1 + \cos(n_{ijkl} \chi_{ijkl} - \delta_{ijkl})], \quad (12)$$

where  $k_{ijkl}^d$ ,  $n_{ijkl}$  and  $\delta_{ijkl}$  are parameters of the potential, and  $\chi_{ijkl}$  is the angle between the planes formed by atoms  $i, j, k$  and  $j, k, l$ .

The dihedral interactions also become disturbed upon covalent bond rupture; therefore, Eq. (12) should be modified to properly account for this effect. The rupture of a dihedral interaction between a quadruple of atoms  $i, j, k$  and  $l$  should take into account three bonds that contribute to this interaction. Thus, the potential energy describing the dihedral interaction with account for the bond rupture reads as:

$$\tilde{U}_{ijkl}^{(\text{dihedral})} = \sigma(r_{ij}) \sigma(r_{jk}) \sigma(r_{kl}) U_{ijkl}^{(\text{dihedral})}, \quad (13)$$

where  $U_{ijkl}^{(\text{dihedral})}$  is the potential (12) describing the dihedral interaction within the framework of

the standard CHARMM force field. The functions  $\sigma(r_{ij})$ ,  $\sigma(r_{jk})$ , and  $\sigma(r_{kl})$  are defined by Eq. (10); they are used to limit the dihedral interaction upon increasing the corresponding bond length. Figure 4b shows a typical profile of a dihedral potential with accounting for the bond rupture. In this case, we have considered the CN4 – P – ON2 – CN7 dihedral interaction where the middle ON2 – P bond was broken. This interaction is also important when modeling bond breakages in DNA nucleotides.

## 2.5 Formation of new bonds

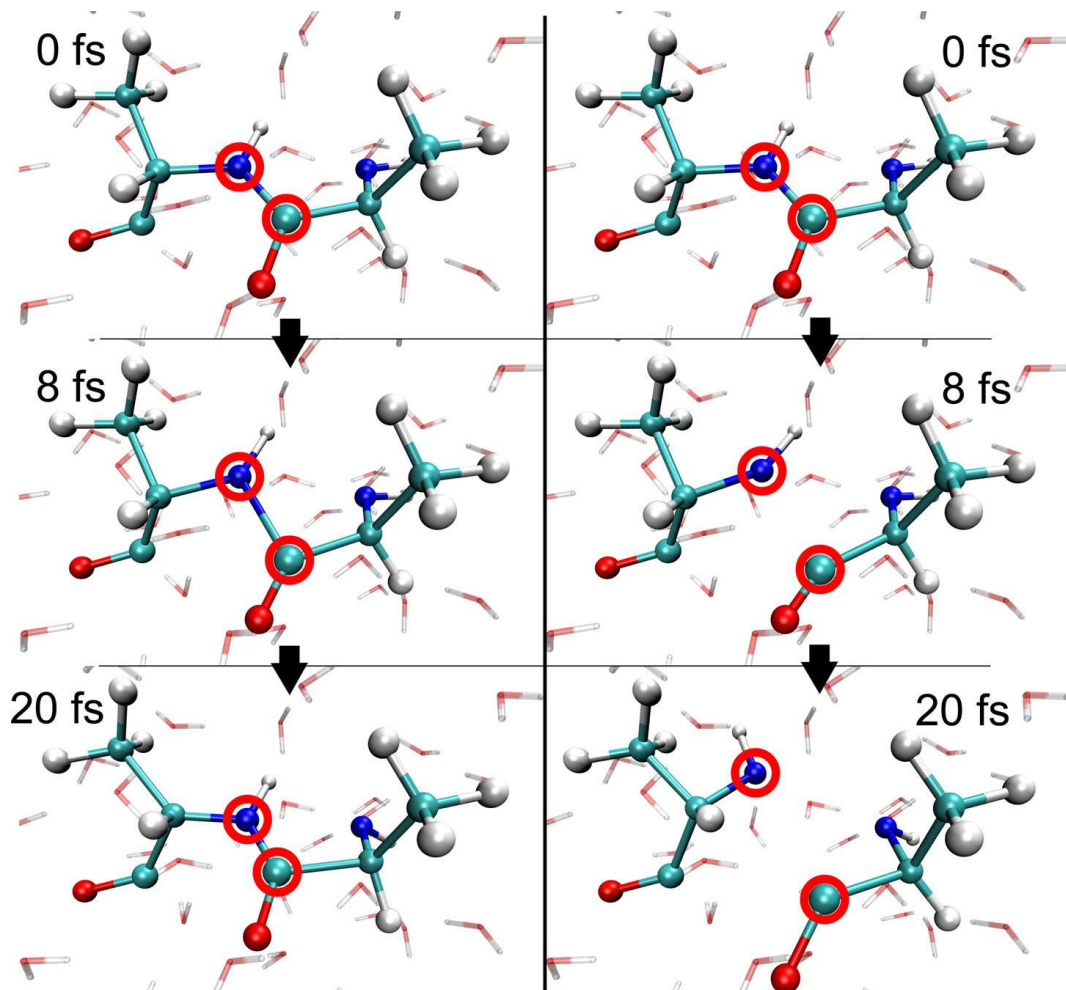
Rupture of covalent bonds leads to formation of individual atoms, radicals or smaller molecular fragments. In order to properly simulate the chemical balance in the system, one should allow for the formation of new bonds.

In MBN EXPLORER, after the rupture of a chemical bond between two atoms, these atoms are placed in a special list of chemically active atoms. Only the atoms from this list can participate in chemical reactions and form new bonds. For each atom from this list, the number of possible molecular bonds is stored and determined by its valence.

In order to create new bonds in the system, the list of chemically active atoms is evaluated at each simulation step, and the neighboring atoms are selected. A chemical bond is formed between a pair of atoms provided that the following conditions have been met: (i) the parameters of the bond, e.g., the equilibrium distance and the value of bond formation energy for this combination of atoms are defined in the simulation input, (ii) atoms are modeled as bound through the Morse potential, and (iii) the distance between the atoms is less than the predefined cutoff/capture radius, which is an independent parameter for each bond type used to speed-up the simulations. If all these conditions are met simultaneously in the simulation, the bond is created and the system's topology is updated. For each new covalent bond, the neighboring atoms are analyzed. If the parameters of angular bond are defined for some group of atoms such bond should also be formed.

## 2.6 Redistribution of partial charges

The rupture and formation of bonds leads to re-configuration of molecular systems and to redistribution of partial charges of atoms. The modification of MBN EXPLORER presented in this paper accounts for this phenomenon. The charge redistribution should obey the following conditions: the total charge of the systems is conserved, the total charge of each individual molecule is an integer in atomic system of units, i.e. an integer number of the electron charge. MBN EXPLORER supports the



**Fig. 5.** Snapshots illustrating dynamics of alanine dipeptide and the C-N bond rupture simulated with the harmonic (left) and Morse (right) potentials at 0 fs (top), 8 fs (middle) and 20 fs (bottom).

two options for modeling the charge redistribution process: (i) a general (default) one applicable to any molecule, and (ii) a special one where charges within molecules are redistributed according to the known electronic configurations.

The default mechanism of charge redistribution is activated upon rupture or formation of covalent bonds. In the case of a bond rupture, two newly emerged fragments of a molecular system have likely non-compensated, non-integer charges. The total charge of each of the newly emerged fragment is thus rounded to the closest integer value and the charge difference is transferred from one fragment to another. This difference is redistributed evenly among all atoms of the fragments. Upon the formation of new a bond, the charge is redistributed inside the newly created molecule in order to lower the values of partial charges preserving the initial sum of charges.

### 3 Numerical Results

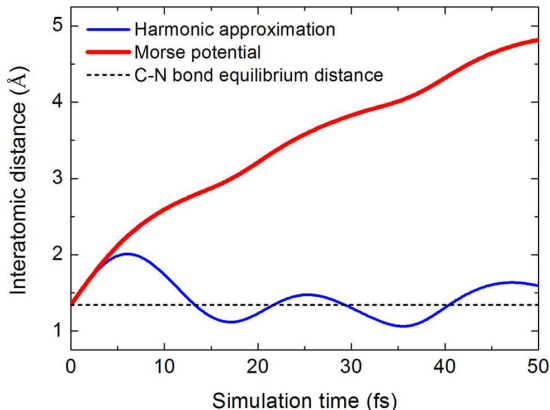
Let us now consider the case studies that go beyond the standard MM methodology. The first example illustrates the rupture of a single C – N bond in an alanine dipeptide molecule. In the second example, we simulate the reverse process of the new bond formation. These two case studies demonstrate a proof of principle for the modified MM force field to describe the dynamically changing molecular topology of the system within the classical MD framework. The third example is devoted to the investigation of the process of water splitting at high temperatures. In this case, the breakage of chemical bonds and the formation of new ones lead to establishing the chemical equilibrium in the system.

#### 3.1 Fragmentation of alanine dipeptide

To illustrate the bond breakage, we have simulated the dynamics of alanine dipeptide consisting of 20

atoms, solvated in a simulation box with 95 water molecules. The alanine dipeptide molecule was considered with neutral terminals.

In order to clearly illustrate the difference between the standard CHARMM force field, utilizing the harmonic interatomic potential, and the dissociative CHARMM potential implemented in MBN EXPLORER, two simulations were carried out. In these simulations, the rupture of the central C–N bond in the dipeptide, leading to the formation of two isolated alanine molecules, was monitored. To facilitate the process, we have set the initial velocity of the C and N atoms high, which corresponds to an energy fluctuation sufficient for the bond rupture. In the simulation performed with the standard force field, the peptide bond is modeled through the harmonic potential, therefore, the bond cannot break. The behavior of the C–N bond in the harmonic approximation is illustrated in the left part of Figure 5, and the corresponding atoms are marked with red circles. In this case, the distance between the atoms oscillates around the equilibrium value as the atoms always return to their equilibrium positions.



**Fig. 6.** Dependence of the C–N interatomic distance in alanine dipeptide as a function of the simulation time.

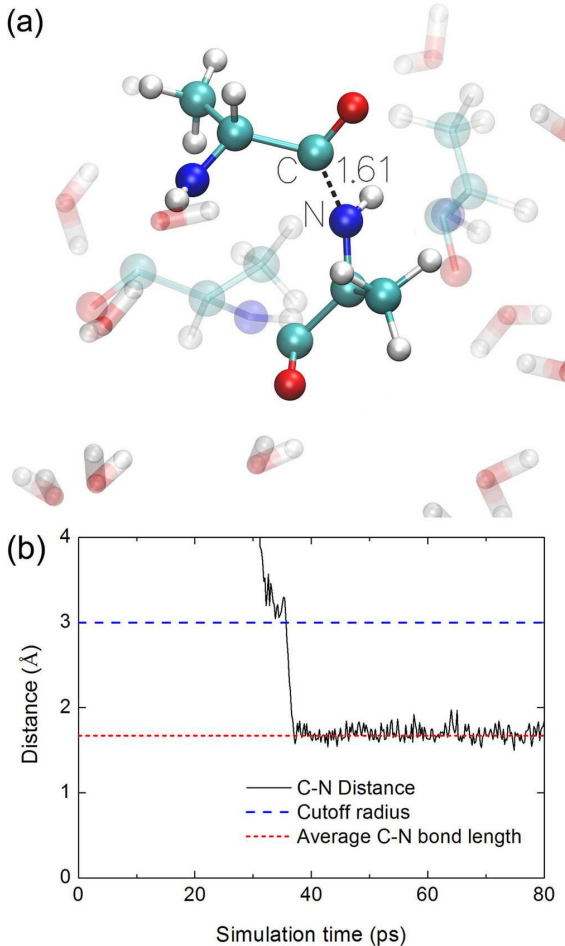
In the second simulation, the Morse potential (5) was used for the description of the peptide bond. In this case the C and N atoms do not oscillate around an equilibrium position, and the structure of the system after 20 fs of simulation changes significantly from the one considered above (see the right part of Figure 5). It is evident from the snapshots that the distance between the atoms increases already after 8 fs.

When the distance between the atoms exceeds a given cutoff radius (which is equal to 2.5 Å in this example), the bond is considered as broken. Once this has happened, the carbon and the nitrogen atoms remain interacting only via the electrostatic potential and the van der Waals interactions,

so that the two alanine molecules can diffuse apart. The charge redistribution does not happen in this case because both new fragments of the dipeptide were initially neutral.

Figure 6 shows the interatomic distance between the carbon and the nitrogen atoms as a function of the simulation time. The equilibrium distance between the atoms is  $r_0^{C-N} = 1.354$  Å (dashed line). The figure demonstrates that in the case of the simulation with the Morse potential, the interatomic distance monotonically increases indicating that the bond is broken and that two isolated alanine molecules drift apart.

### 3.2 Binding of two alanine amino acids



**Fig. 7.** (a) Two alanine molecules approaching each other to form a new C–N bond. (b) Dependence of the distance between C and N atoms for the two alanine molecules.

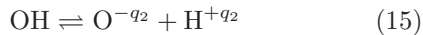
The second example illustrates the process of binding two alanine molecules together into a single dipeptide through the formation of a new covalent

bond in the molecular system. In this case study, six isolated alanine amino acids surrounded by 54 water molecules were placed in a small simulation box of  $24 \times 24 \times 24 \text{ \AA}^3$  with periodic boundary conditions, and the dynamics of the system was simulated for 80 ps at a fixed temperature of 1000 K controlled by the Langevin thermostat with the damping time constant of 1 fs. Each alanine molecule was modeled with unsaturated N- and C- termini, i.e. having two unpaired chemical bonds.

In the course of the simulation, all distances between the different termini of alanines were monitored. When the distance between a pair of terminal atoms became smaller than the predefined cutoff radius (equal to 3 Å in this example), a new covalent bond was considered to be formed. Figure 7a illustrates a spacial conformation of two amino acids in the simulation leading to the formation of a new bond. Figure 7b gives the dependence of the distance between C and N atoms for the two molecules shown in the upper part. At some point, this distance becomes smaller than the cutoff radius (blue dashed line), and the two molecules become connected. Note that after 40 ps the distance between the C and N atoms oscillates around a constant value corresponding to the C-N bond equilibrium length. Since six alanines are considered in this simulation, more of them could self-assemble in a polypeptide chain but this would require longer simulation. In this system, each initial alanine molecule has a total charge equal to zero. Therefore, after the formation of a new molecule the charge redistribution step was not necessary.

### 3.3 Water splitting

The third case study concerns the process of water splitting. At elevated temperatures, water molecules can dissociate forming different molecular products. In such a system, the following reactions take place:



which describe both the mechanisms of rupture and formation of new chemical bonds. The superscripts  $+q_i$  and  $-q_i$  ( $i = 1, 2$ ) indicate that the reaction products in Eqs. (14) and (15) carry some non-zero charge. The principle of detailed balance states that for a reversible process at the equilibrium a direct process should be equilibrated by its reverse process. This forms the equilibrium concentrations of reacting species which depend on the temperature of the medium. In order to determine the equilibrium concentrations of molecules in our system,

one should use the equations for the equilibrium concentrations of different molecular species [59], which are based on the Dalton's law and the law of mass action [60, 61]:

$$\frac{X_{\text{H}}X_{\text{OH}}P}{X_{\text{H}_2\text{O}}} = \exp\left(-\frac{\Delta G_1}{RT}\right) \quad (18)$$

$$\frac{X_{\text{H}}X_{\text{O}}P}{X_{\text{OH}}} = \exp\left(-\frac{\Delta G_2}{RT}\right) \quad (19)$$

$$\frac{X_{\text{H}}^2P}{X_{\text{H}_2}} = \exp\left(-\frac{\Delta G_3}{RT}\right) \quad (20)$$

$$\frac{X_{\text{O}}^2P}{X_{\text{O}_2}} = \exp\left(-\frac{\Delta G_4}{RT}\right) , \quad (21)$$

where  $X_{\text{H}}$ ,  $X_{\text{OH}}$ ,  $X_{\text{O}}$  and  $X_{\text{H}_2\text{O}}$  are the relative concentrations of atoms and molecules of different types noted by the subscripts,  $\Delta G_i$  are Gibbs energy values – the dissociation energies of the chemical reactions in Eqs. (14)-(17),  $P$  is the total pressure in the system,  $R$  is the universal gas constant, and  $T$  is the temperature. Taking into account that the total number of hydrogen atoms in the studied system is twice the number of oxygen atoms, one obtains:

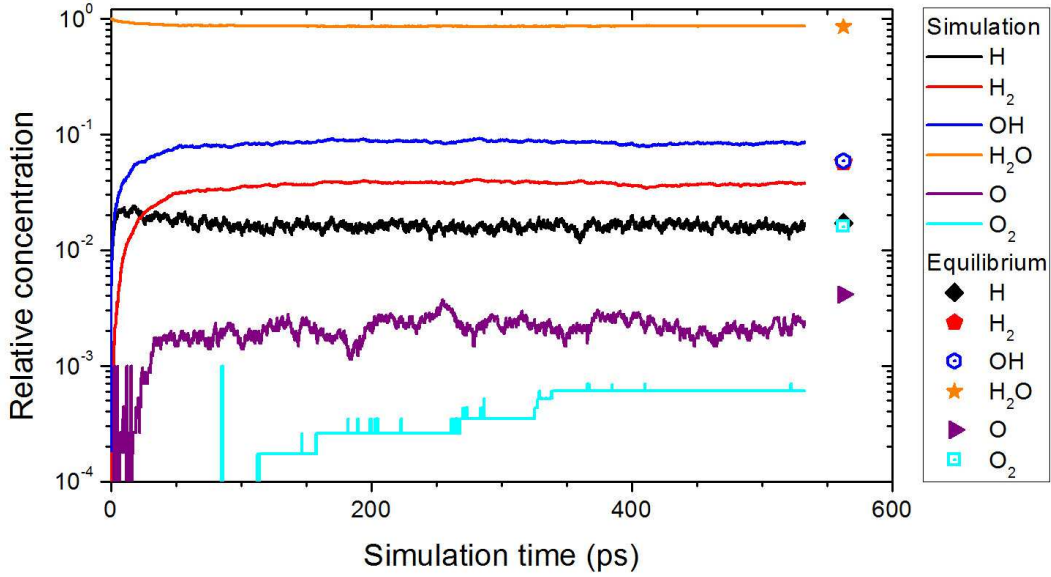
$$\begin{aligned} & 2X_{\text{H}_2} + X_{\text{H}} + X_{\text{OH}} + 2X_{\text{H}_2\text{O}} \\ & = 2(2X_{\text{O}_2} + X_{\text{OH}} + X_{\text{O}} + X_{\text{H}_2\text{O}}) . \end{aligned} \quad (22)$$

The normalization condition for the partial concentrations/densities can be written accordingly as:

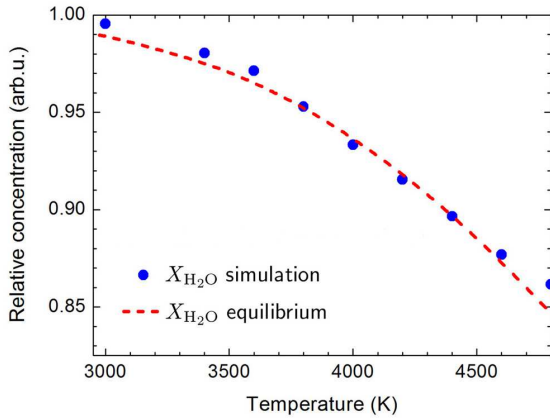
$$\begin{aligned} & X_{\text{H}_2} + X_{\text{O}_2} + X_{\text{H}} \\ & + X_{\text{OH}} + X_{\text{O}} + X_{\text{H}_2\text{O}} = 1 . \end{aligned} \quad (23)$$

In this case study, we have simulated water splitting in a box of  $130 \times 130 \times 130 \text{ \AA}^3$  with periodic boundary conditions. The studied system consists of 10917 water molecules randomly placed inside this box. Simulations for different temperatures were carried out using the Langevin thermostat with the damping time of 10 fs. The total simulation time was set to 500 ps. We note that accounting for the dissociation and formation of new bonds in the MM method with dynamic topology does not affect significantly the performance of the simulation. The computational cost of the calculation is about the same as for the conventional MD simulation with the standard CHARMM force field; the simulation of water splitting was carried out using a 12-core AMD workstation in 3 days. Due to the constant volume and density, and different temperatures at each simulation, the pressure in the box was changing from 2500 to 4000 bar which corresponds to the water vapor state with a high density. The simulated values of pressure and temperature were used in the analytic expressions, Eqs. (18)-(23), for estimating the equilibrium concentrations of the reaction products. The values of





**Fig. 8.** Time evolution of the relative concentration of atoms  $X_H$ ,  $X_O$  and molecules  $X_{H_2}$ ,  $X_{OH}$ ,  $X_{H_2O}$  in the simulated system at  $T = 4800$  K. Results of the simulation are shown with solid lines, dots show corresponding analytic values of concentrations.



**Fig. 9.** Comparison of the simulated dependence of the relative concentration of water molecules,  $X_{H_2O}$ , on temperature with the one obtained from the equilibrium analysis, Eqs. (18)-(23).

$G_1 = G_2 = 117.9$  kcal/mol,  $G_3 = 104$  kcal/mol,  $G_4 = 119$  kcal/mol were used.

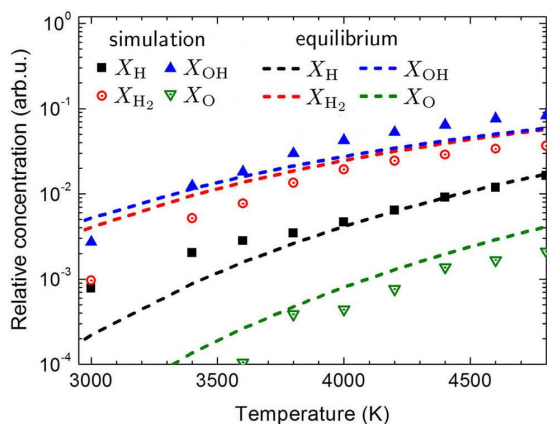
For each reaction (14)–(15), the distribution of partial charges  $q_i$  was calculated separately and set explicitly for all products. For water, the partial charge of oxygen and hydrogen atoms was set to  $-0.834$  and  $0.417$ , respectively, for a OH molecule the values of  $-0.375$  and  $0.375$  were assumed.  $H_2$  and  $O_2$  molecules in reactions (16) and (17) were considered as neutral with zero partial charges.

Figure 8 shows the time evolution of the relative concentration of atoms and molecules in the system at  $T = 4800$  K. Solid lines present the results of the simulations, while symbols show the results of numerical solution of Eqs. (18)–(23).

Initially, all molecules in the system are water molecules. Frequent random collisions lead to their dissociation into H and OH. During the first 10 ps, these are the two main reaction products and their relative concentration coincide. After 10 ps the concentration of hydrogen atoms becomes significant and the reaction  $H + H \rightarrow H_2$  becomes probable. At this instance the concentration of  $H_2$  increases significantly and the concentration of H drops. After 100 ps of simulation, the concentrations of H,  $H_2$ , OH and  $H_2O$  become nearly constant and close to the equilibrium values obtained from Eqs. (18)–(23). However due to the low concentration of oxygen atoms the formation of  $O_2$  molecules is a relatively slow process and requires significant time which we could not reach in our simulation. This process starts after about 100 ps and does not reach the saturation stage by the end of the simulation.

To further characterize the interconversion of chemical reactions in the system, in Figure 9 we present the comparison of the simulated temperature dependence of the relative concentration of water molecules,  $X_{H_2O}$ , with the one obtained from the equilibrium analysis, Eqs. (18)–(23). The increase of temperature leads to the splitting of water molecules and their number decreases. The comparison shows that the two curves show practically identical behavior. Small deviations can be attributed to the limited simulation time and to the high pressure in the system, as both of these factors lead to the decrease of the number of split molecules.

A similar analysis was performed for  $X_H$ ,  $X_{H_2}$ ,  $X_{OH}$ , and  $X_O$ . Figure 10 illustrates the dependence of relative concentration on temperature for the different products of dissociation. These dependencies



**Fig. 10.** Comparison of simulated dependencies of  $X_H$ ,  $X_{H_2}$ ,  $X_{OH}$ ,  $X_O$  on temperature with those obtained from the equilibrium analysis, Eqs. (18)-(23). Dashed lines correspond to the results of the equilibrium analysis, solid curves correspond to the results of simulations.

are compared with those obtained from the equilibrium analysis based on Eqs. (18)-(23). The figure demonstrates a reasonable agreement of the two approaches, especially at high temperatures where the simulated values  $X_H$  nearly coincide with those obtained from the equilibrium analysis, and the values of  $X_{H_2}$  and  $X_{OH}$  are rather close. For smaller temperatures, the deviation of the curves increases, primarily because of the insufficient simulation time. At higher temperatures, the molecules diffuse faster, thus allowing to achieve sooner the chemical equilibrium in the system.

## 4 Conclusions

This paper reports on an important development of the MBN EXPLORER software package, namely on the implementation of reactive force fields based on the modified CHARMM force field. This modification allows one to simulate and analyze chemical reactions and transformations involving covalent bond breakages in bio- and organic molecules and molecular systems [62, 63]. This is obviously a very large research domain and we believe that the implementations reported in this work will open many new possibilities for the research analysis based on the computational modeling of the mentioned systems. In this paper, we have not tried to explore all the possible applications of the new computational tool, but instead focused on the most characteristic and illustrative examples.

This work can be extended further in many different ways. The introduced reactive force fields can be applied systematically to different systems and the results of simulations can be compared with the appropriate experiments [62, 63]. The simulation

results can also be validated through the comparison with the corresponding results obtained through the quantum MD simulations of relatively small molecular systems [39, 40]. These and many more other possible studies based on the reported key implementation are the subject for further research.

## Acknowledgments

The possibility to perform computer simulations at the Frankfurt Center for Scientific Computing is gratefully acknowledged. The authors acknowledge supercomputer time on Stampede provided the Texas Advanced Computing Center (TACC) at the University of Texas at Austin through Extreme Science and Engineering Discovery Environment (XSEDE) Grant XSEDE MCB-120160 (to IAS). IAS is grateful for the financial support from the Lundbeck Foundation and to the Russian Scientific Foundation (Grant No. 14-12-00342).

## References

1. K.Y. Sanbonmatsu, C.S. Tung, J. Struct. Biol. **157**, 470 (2007)
2. G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken et al., Nature **497**, 643 (2013)
3. R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, J. Chem. Theory Comput. **9**, 3878 (2013)
4. A.K. Rappé, C.J. Casewit, *Molecular Mechanics Across Chemistry* (University Science Books, 1997)
5. J. Gumbart, E. Schreiner, D.N. Wilson, R. Beckmann, K. Schulten, Biophys. J. **103**, 331 (2012)
6. J. Shim, G.I. Humphreys, B.M. Venkatesan, J.M. Munz, X. Zou, C. Sathe, K. Schulten, F. Kosari, A.M. Nardulli, G. Vasmatazis et al., Sci. Rep. **3**, 1389 (2013)
7. B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, J. Comput. Chem. **4**, 187 (1983)
8. D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, J. Comput. Chem. **26**, 1668 (2005)
9. D.V.D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, H. Berendsen, J. Comput. Chem. **26**, 1701 (2005)
10. J.C. Phillips, R. Braun, W. Wang, et al, J. Comput. Chem. **26**, 1781 (2005)
11. I.A. Solov'yov, A.V. Yakubovich, P.V. Nikolaev, I. Volkovets, A.V. Solov'yov, J. Comput. Chem. **33**, 2412 (2012)

12. V.V. Dick, I.A. Solov'yov, A.V. Solov'yov, J. Phys.: Conf. Ser. **248**, 012025 (2010)
13. V.V. Dick, I.A. Solov'yov, A.V. Solov'yov, Phys. Rev. B **84**, 115408 (2011)
14. M. Panshenskov, I.A. Solov'yov, A.V. Solov'yov, J. Comput. Chem. **35**, 1317 (2014)
15. I.A. Solov'yov, A.V. Solov'yov, N. Kébaili, A. Masson, C. Bréchnignac, Phys. Status Solidi B **251**, 609 (2014)
16. I.A. Solov'yov, A.V. Solov'yov, J. Phys.: Conf. Ser. **438**, 012006 (2013)
17. G.B. Sushko, V.G. Bezchastnov, I.A. Solov'yov, A.V. Korol, W. Greiner, A.V. Solov'yov, J. Comput. Phys. **252**, 404 (2013)
18. G.B. Sushko, A.V. Korol, W. Greiner, A.V. Solov'yov, J. Phys.: Conf. Ser. **438**, 012018 (2013)
19. G.B. Sushko, V.G. Bezchastnov, A.V. Korol, W. Greiner, A.V. Solov'yov, R.G. Polozkov, V.K. Ivanov, J. Phys.: Conf. Ser. **438**, 012019 (2013)
20. G.B. Sushko, A.V. Korol, A.V. Solov'yov, Nucl. Instrum. Meth. B **355**, 39 (2015)
21. A.V. Verkhovtsev, S. Schramm, A.V. Solov'yov, Eur. Phys. J. D **68**, 246 (2014)
22. A.V. Verkhovtsev, M. Hanauske, A.V. Yakubovich, A.V. Solov'yov, Comput. Mater. Sci. **76**, 80 (2013)
23. G.B. Sushko, A.V. Verkhovtsev, A.V. Solov'yov, J. Phys. Chem. A **118**, 8426 (2014)
24. I.A. Solov'yov, M. Mathew, A.V. Solov'yov, W. Greiner, Phys. Rev. E **78**, 051601 (2008)
25. I.A. Solov'yov, J. Geng, A.V. Solov'yov, B.F.G. Johnson, Chem. Phys. Lett. **472**, 166 (2009)
26. J. Geng, I.A. Solov'yov, W. Zhou, A.V. Solov'yov, B.F.G. Johnson, J. Phys. Chem. C **113**, 6390 (2009)
27. J. Geng, I.A. Solov'yov, D.G. Reid, P. Skelton, A.E. Wheatley, A.V. Solov'yov, B.F.G. Johnson, Phys. Rev. B **81**, 214114 (2010)
28. P. Moskovkin, M. Panshenskov, S. Lucas, A.V. Solov'yov, Phys. Status Solidi B **251**, 1456 (2014)
29. G.B. Sushko, A.V. Verkhovtsev, A.V. Yakubovich, S. Schramm, A.V. Solov'yov, J. Phys. Chem. A **118**, 6685 (2014)
30. A.V. Verkhovtsev, A.V. Yakubovich, G.B. Sushko, M. Hanauske, A.V. Solov'yov, Comput. Mater. Sci. **76**, 20 (2013)
31. A.V. Yakubovich, E. Surdutovich, A.V. Solov'yov, J. Phys.: Conf. Ser. **373**, 012014 (2012)
32. E. Surdutovich, A.V. Yakubovich, A.V. Solov'yov, Sci. Rep. **3**, 1289 (2013)
33. E. Surdutovich, A.V. Solov'yov, Eur. Phys. J. D **68**, 353 (2014)
34. A.V. Solov'yov, E. Surdutovich, E. Scifoni, I. Mishustin, W. Greiner, Phys. Rev. E **79**, 011909 (2009)
35. E. Surdutovich, A.V. Yakubovich, A.V. Solov'yov, Eur. Phys. J. D **60**, 101 (2010)
36. H.M. Senn, W. Thiel, Angew. Chem. Int. Ed. **48**, 1198 (2009)
37. R. Car, M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)
38. D. Marx, J. Hutter, *Ab initio Molecular Dynamics: Theory and Implementation*, in *Modern Methods and Algorithms of Quantum Chemistry* (John von Neumann Institute for Computing, Jülich, Germany, 2000), p. 301
39. M. Smyth, J. Kohanoff, J. Am. Chem. Soc. **134**, 9122 (2012)
40. M. McAllister, M. Smyth, B. Gu, G.A. Tribello, J. Kohanoff, J. Phys. Chem. Lett. **6**, 3091 (2015)
41. G. Lüdemann, I.A. Solov'yov, T. Kubař, M. Elstner, J. Am. Chem. Soc. **137**, 1147 (2015)
42. E. Sjulstok, J.M. Haugaard Olsen, I.A. Solov'yov, Sci. Rep. (accepted)
43. D.A. Beck, R.S. Armen, V. Daggett, Biochemistry **44**, 609 (2005)
44. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **46**, 227 (2008)
45. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, *On the Theory of Phase Transitions in Polypeptides*, in *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale* (World Scientific, 2008), Vol. 1, p. 241
46. A.V. Yakubovich, I. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **51**, 25 (2009)
47. S. Piana, J.L. Klepeis, D.E. Shaw, Curr. Opin. Struct. Biol. **24**, 98 (2014)
48. E.S. Henriques, A.V. Solov'yov, Eur. Phys. J. D **46**, 471 (2008)
49. I.A. Solov'yov, T. Domratcheva, A.R. Moughal Shahi, K. Schulten, J. Am. Chem. Soc. **134**, 18046 (2012)
50. I.A. Solov'yov, T. Domratcheva, K. Schulten, Sci. Rep. **4**, 3845 (2014)
51. A.M. Barragan, A.R. Crofts, K. Schulten, I.A. Solov'yov, J. Phys. Chem. B **119**, 433 (2014)
52. K. Hart, N. Foloppe, C.M. Baker, E.J. Denning, L. Nilsson, A.D. MacKerell Jr, J. Chem. Theory Comput. **8**, 348 (2011)
53. X. Zou, W. Ma, I.A. Solov'yov, C. Chipot, K. Schulten, Nucleic Acids Res. **40**, 2747 (2012)
54. S.N. Volkov, E.V. Paramonova, A.V. Yakubovich, A.V. Solov'yov, J. Phys.: Condens. Matter **24**, 035104 (2012)
55. J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell Jr, R.W.

- Pastor, J. Phys. Chem. B **114**, 7830 (2010)
56. R. Pastor, A. MacKerell Jr, J. Phys. Chem. Lett. **2**, 1526 (2011)
  57. Å.A. Skjevik, B.D. Madej, C.J. Dickson, K. Teigen, R.C. Walker, I.R. Gould, Chem. Commun. **51**, 4402 (2015)
  58. A.D. MacKerell, D. Bashford, R.L. Bellott, *et al*, J. Phys. Chem. B **102**, 3586 (1998)
  59. S. Ihara, Int. J. Hydrogen Energy **3**, 287 (1978)
  60. L. Landau, E. Lifshitz, *Statistical Physics*, 3rd. ed., Course of Theoretical Physics Vol. 5 (Butterworth-Heinemann, 1980)
  61. P. Atkins, J. De Paula, V. Walters, *Physical Chemistry* (Macmillan Higher Education, 2006), ISBN 9780716774334
  62. J. de Vries, R. Hoekstra, R. Morgenstern, T. Schlathölter, J. Phys. B: At. Mol. Opt. Phys. **35**, 4373 (2002)
  63. C. Zhou, S. Matsika, M. Kotur, T.C. Weinacht, J. Phys. Chem. A **116**, 9217 (2012)